Year: 2020

A priori power considerations in orthodontic research: a 3 year meta-epidemiologic study

Gratsia, Sophia ; Koletsi, Despina ; Fleming, Padhraig S ; Pandis, Nikolaos

Abstract: AIM: To assess the prevalence of a priori power calculations in orthodontic literature and to identify potential associations with a number of study characteristics, including journal, year of publication and statistical significance of the outcome. MATERIALS AND METHODS: The electronic archives of four leading orthodontic journals with the highest impact factor (American Journal of Orthodontics and Dentofacial Orthopedics, AJODO; European Journal of Orthodontics, EJO; Angle Orthodontist, ANGLE; Orthodontics and Craniofacial Research, OCR) were assessed over a 3 year period until December 2018. The proportion of articles reporting a priori power calculations were recorded, and the association with journal, year of publication, study design, continent of authorship, number of centres and researchers, statistical significance of results and reporting of confidence intervals (CIs) was assessed. Univariable and multivariable regression were used to identify significant predictors. RESULTS: Overall, 654 eligible articles were retrieved, with the majority published in the AJODO (n = 246, 37.6%), followed by ANGLE (n = 222, 33.9%) and EJO (n = 139, 21.3%). A total of 233 studies (35.6%) presented power considerations a priori along with sample size calculations. Study design was a very strong predictor with interventional design presenting 3.02 times higher odds for a priori power assumptions compared to observational research [odds ratio (OR): 3.02; 95% CIs: 2.06, 4.42; P < 0.001]. CONCLUSIONS: Presentation of a priori power considerations for sample size calculations was not universal in contemporary orthodontic literature, while specific study designs such as observational or animal and in vitro studies were less likely to report such considerations.

DOI: https://doi.org/10.1093/ejo/cjz061

Originally published at:
Gratsia, Sophia; Koletsi, Despina; Fleming, Padhraig S; Pandis, Nikolaos (2020). A priori power considerations in orthodontic research: a 3 year meta-epidemiologic study. European Journal of Orthodontics, 42(4):454-459.
DOI: https://doi.org/10.1093/ejo/cjz061

Title: A priori power considerations in orthodontic research: a 3- year meta-epidemiologic study.

Sophia Gratsia[1], Despina Koletsi[2], Padhraig S. Fleming[3], Nikolaos Pandis[4]

[1] School of Dentistry, National and Kapodistrian University of Athens, Athens, Greece

[2] Department of Orthodontics, School of Dentistry, National and Kapodistrian University of Athens, Athens, Greece and Clinic of Orthodontics and Paediatric Dentistry, Center of Dental Medicine, University of Zurich, Zurich, Switzerland

[3] Centre for Oral Bioengineering, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

[4] Department of Orthodontics and Dentofacial Orthopedics, Dental School/Medical Faculty, University of Bern, Switzerland and Private Practice in Corfu, Greece

Publication:

Gratsia S, Koletsi D, Fleming PS, Pandis N. A priori power considerations in orthodontic research: a 3- year meta-epidemiologic study. Eur J Orthod. 2019 Aug 15. pii: cjz061. doi: 10.1093/ejo/cjz061. [Epub ahead of print].

Short title: Power calculations in orthodontic research

Correspondence: Dr. Despina Koletsi, 2 Thivon St, 11527 Goudi, Attica, Greece; email: d.koletsi@gmail.com, dkoletsi@dent.uoa.gr

Summary

Aim: To assess the prevalence of *a priori* power calculations in orthodontic literature and to identify potential associations with a number of study characteristics including journal, year of publication and statistical significance of the outcome.

Materials and Method: The electronic archives of four leading orthodontic journals with the highest impact factor (American Journal of Orthodontics and Dentofacial Orthopedics; AJODO, European Journal of Orthodontics; EJO, Angle Orthodontist; ANGLE, Orthodontics and Craniofacial Research; OCR) were assessed over a 3-year period until December 2018. The proportion of articles reporting *a priori* power calculations were recorded, and the association with journal, year of publication, study design, continent of authorship, number of centers and researchers, statistical significance of results, and reporting of confidence intervals was assessed. Univariable and multivariable regression were used to identify significant predictors.

Results: Overall, 654 eligible articles were retrieved with the majority published in the AJODO (n=246, 37.6%), followed by ANGLE (n=222, 33.9%) and EJO (n=139, 21.3%). A total of 233 studies (35.6%) presented power considerations *a priori* along with sample size calculations. Study design was a very strong predictor with interventional design presenting 3.02 times higher odds for *a priori* power assumptions compared to observational research (OR: 3.02, 95% CIs: 2.06, 4.42; $p<0.001$).

Conclusions: Presentation of *a priori* power considerations for sample size calculations was not universal in contemporary orthodontic literature, while specific study designs such as observational or animal and *in-vitro* studies were less likely to report such considerations.

Keywords: statistical power, power calculation, sample size, orthodontic

Introduction

Methological and reporting flaws are prevalent among medical and dental studies, and orthodontic research Is not immune to pitfalls related to design, conduct and reporting (1-3). Although compliance with reporting guidelines has been actively endorsed by journal editors in an attempt to promote clinical decision-making based on correct inferences and interpretation of research findings (4,5), unclear and suboptimal reporting persists (6-9).

Sample size calculation is imperative (10) when designing a study or clinical trial and is often recommended in other study designs. Having inadequate participant numbers in a clinical trial is likely to yield inconclusive results leading to research waste, whereas including more patients than required may expose patients unnecessarily to ineffective or potentially harmful treatments. The importance of power calculations is emphasized in reporting guidelines such as the CONSORT statement for randomized controlled trials (11-13). The power of a study is defined as *1- beta*, where *beta* is the risk of type II error (false negative) and refers to the probability of not observing a statistically significant difference when one actually exists. In other words, study power indicates the probability of identifying a difference when such a difference truly exists (14).

It has been common practice across several biomedical fields to advocate *post- experiment* or *post-hoc* power calculations to justify statistically non-significant findings on the basis of constraints concerning recruitment of manageable or convenient number of subjects (15). However, such practices have been associated with erroneous inferences and interpretation of research findings (16). Inability to pre-define the desired power of a study is likely to deter researchers from obtaining the optimal sample size to detect a statistically and clinically important treatment effect, risking research waste in terms of cost and time (17).

Therefore, we aimed to assess the prevalence of reporting of *a priori* power calculations within contemporary orthodontic literature and to identify the potential effect of a number of variables on this practice such as journal, study design or timing of publication over a 3-year period.

Materials and methods

The contents of 4 major orthodontic journals with the highest impact factor were electronically searched over a period of 3 years and until December 2018, to identify publications that could potentially present either *a priori* or *post-hoc* power calculations. Journals assessed included: American Journal of Orthodontics and Dentofacial Orthopedics (AJODO), Angle Orthodontist (ANGLE), European Journal of Orthodontics (EJO), and Orthodontics and Craniofacial Research (OCR). The entire list of original studies was considered eligible for inclusion, excluding editorials, case reports, cross-sectional studies, pilot studies, opinion letters and reviews. Included studies were classified based on their design as interventional or observational in human subjects, while laboratory or animal studies were also considered separately.

Data extraction was carried out based on pre-specified standardized piloted forms and calibration between the two assessors (SG, DK) was undertaken on 30 articles. Inter-rater agreement between examiners was assessed on 20 additional papers. Whether studies included *a priori* or *post-hoc* power calculations was the primary outcome assessed. Additionally, a number of characteristics and predictor variables were examined: journal, year of publication, study design, geographic region denoting affiliation of the first author, number of centers (single or multi-center, based on affiliation details and additional details about the place of the study within the Materials and Methods section), number of researchers participating in the publication, whether the primary outcome pertained to a statistically significant effect and whether Confidence Intervals (CIs) were reported.

Statistical analysis

Descriptive statistics were performed for the pre-defined variables. Cross-tabulations were constructed to test the association between reporting of *a priori* power calculations or otherwise and study characteristics, through chi- square tests. Univariable and multivariable logistic regression models were implemented to examine the effect of study characteristics including journal, year of publication and study design on conducting of *a priori* power calculations. The predictors were examined sequentially one at a time in the initial model and retained in the final multivariable model if $p < 0.10$. In addition, journal was considered an *a priori* predictor and was retained in the final model. The Hosmer-Lemeshow test was used to check model fit. The unweighted kappa statistic was used to assess inter-rater agreement on the reported power calculations. The predefined level of significance was set at $p<0.05$. All analyses were conducted with Stata version 15.1 (Stata Corporation, College Station, Texas, USA).

Results

A total of 982 articles were screened within the 3-year period, of which 654 were suitable for inclusion (Figure 1, Appendix 1). Inter-rater agreement yielded an unweighted kappa statistic of 0.86 indicating almost perfect agreement between the two investigators (i.e. recording of *a priori* power considerations or otherwise). The highest percentage of the assessed articles were published in the AJODO (246/654, 37.6%), followed closely by ANGLE (222/654, 33.9%) and EJO (139/654, 21.2%). Most articles were published in the years 2016 and 2017 (n=461, 70.5%), originated from Asia/Other (267/654, 40.8%), consisted of multi-center efforts (424/654, 64.8%) and were authored by 4 to 6 researchers (n=406, 62.1%). Observational studies predominated (288/654, 44.0%) followed by interventional designs (205/654, 31.3%). Statistically significant findings for the main outcome were recorded for most of the studies (n=482, 73.7%), while

Confidence Intervals (CIs) for the estimated effect were only reported in 119 studies (18.2%) (Table 1). The distribution of statistically significant outcomes or otherwise and reporting of CIs, across different study designs is presented in Table 2.

Overall, almost two-thirds of the studies (421/654, 64.4%) failed to report *a priori* power considerations and either presented post-hoc power calculations (80/421, 19.0%) or nothing at all (341/421, 81.0%). Interventional study design was associated with increased reporting of *a priori* power considerations (119/205, 58%; $p<0.001$). Likewise, presence of non-significant findings for the outcome of interest (78/172, 45.3%; $p=0.002$) and reporting of CIs (57/119, 47.9%; $p=0.002$) were associated with this practice, respectively (Table 1).

According to the multivariable regression model for the effect of article characteristics on the reporting of *a priori* power considerations, there was strong evidence that the study design was a significant predictor of the outcome (p-value for the overall Wald test<0.001), after adjusting for journal, significance of the study findings and reporting of confidence bounds (Figure 2). Specifically, interventional studies presented 3.02 times higher odds for *a priori* considerations compared to observational ones (OR=3.02; 95%CIs: 2.06, 4.42). On the contrary, in vitro research presented 55% lower odds (OR=0.45; 95%CIs: 0.24, 0.83) and animal studies 56% lower odds (OR=0.44; 95%CIs: 0.22, 0.89) for *a priori* considerations compared to observational research. Finally, studies with non-significant findings were associated with 55% higher odds (OR=1.49; 95%CIs: 1.01, 2.20; $p=0.04$) for this practice conditional on journal, study design and reporting of confidence intervals, compared to significant findings (Table 3).

Discussion

The findings of the present empirical study indicate that approximately two-thirds of orthodontic research articles fail to present *a priori* sample size calculations or present *post hoc* calculations.

To the best of our knowledge there is no similar study regarding power considerations across different methodological designs in the existing orthodontic and dental literature and thus no direct comparisons can be attempted either with other dental specialties or as an updated report of the most recent evidence compared to previous knowledge on the topic. Notwithstanding this, there is abundant evidence on the transparency of reporting of sample size calculations in clinical trials within both biomedical (18,11) and dental literature (19,20). It has been claimed that reporting of sample size considerations is suboptimal within randomized controlled trials (RCTs) in orthodontics and other dental areas, with insufficient information presented to allow for sample size recalculations in over 50% thus precluding direct replication of such studies, risking false assumptions and potentially compromising the power of the RCTs (19,20).

Articles stemming from observational or in- vitro/ animal research appeared to be more prone to omitting *a priori* power assumptions compared to interventional studies. This is in keeping with the existing evidence from biomedical literature (6,21) on research practices across these designs. Research conduct and reporting guidelines have been designed across different types of study designs (22,23); however, their adoption by the scientific community, including journal editors, reviewers and investigators, may lag behind that relating to interventional research including RCTs, or indeed as is associated with systematic reviews and meta-analyses of clinical trials.

The preponderance of studies presenting *post-hoc* power calculations is interesting. This statistical practice has been criticised as being misleading and flawed (15). Estimation of high observed power after the completion of an experiment does not translate into stronger evidence for the detected effect. For studies with negative or non-significant effects, the use of *post- hoc* power calculations fail to inform as to whether the observed estimate is a false-negative or a real one. Instead, measures of uncertainty such as confidence bounds have been proposed as a means of estimating the study power post-hoc. As the confidence interval around a point estimate is affected by the sample size of a study, useful information on the estimated treatment effect and its precision is available (24-26).

A higher proportion of studies with *a priori* power estimations were found to report non-significant results for their primary outcome, while reporting of measures of uncertainty was also associated with this desired approach. These findings illustrate that correct practice and accurate conduct and reporting of research may well be followed across several stages of the study design and publication process. Selective reporting and publication of statistically significant results as a common practice has been associated with publication bias (27), while presentation and reporting of non- significant outcomes is equally important. Open and public registration of studies prior to commencement is considered guarantor of clear and transparent reporting. Optimal practices and reporting of one research parameter is likely to be associated with optimal reporting of another, when the same investigators are involved. Moreover, the onus on thorough peer-review processes to expose repeated conduct and reporting issues is clear. Adherence to the existing reporting guidelines seems imperative with an additional training of the scientific community and especially the editors and the reviewers in identifying evidence of research misconduct.

The conclusions of this study are based on a subset of research articles from a finite number of journals. However, this is the first large-scale study on the assessment of power considerations, not only within interventional research and clinical trials, but across different and common study designs. Furthermore, the journals were selected based on their impact on orthodontic readership, while a dynamic and contemporary time span of the most recent publications was selected. As such, the findings are likely to be representative of contemporary research practice within the specialty.

CONCLUSIONS

Based on the findings of the present meta-epidemiologic study, increased awareness of best practice concerning the design of orthodontic studies with *a priori* planned sample size calculations and power considerations should be encouraged. Improved adherence to reporting guidelines is important with researchers requiring awareness of optimal methodological and reporting characteristics.

Conflict of Interest

None to declare.

Funding

None

References

1. Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than P values: estimation rather than hypothesis testing. British medical journal (Clinical research edition), 292(6522), 746-750.
2. Fleming, P.S., Buckley, N., Seehra, J., Polychronopoulou, A. and Pandis, N. (2012) Reporting quality of abstracts of randomized controlled trials published in leading orthodontic journals from 2006 to 2011. American Journal of Orthodontics and Dentofacial Orthopedics, 142(4), 451–458.
3. Koletsi, D., Spineli, L.M., Lempesi, E. and Pandis, N. (2016) Risk of bias and magnitude of effect in orthodontic randomized controlled trials: a meta-epidemiological review. European Journal of Orthodontics, 38(3), 308–312.
4. Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M. and Altman, D.G. (2010) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. British medical journal (Clinical research ed.), 340, c869.
5. Glasziou, P., Meats, E., Heneghan, C. and Shepperd, S. (2008) What is missing from descriptions of treatment in trials and reviews? British medical journal (Clinical research ed.), 336(7659), 1472–1474.
6. Nieuwenhuis, S., Forstmann, B.U. and Wagenmakers, E.J. (2011) Erroneous analyses of interactions in neuroscience: a problem of significance. Nature Neuroscience, 14(9), 1105–1107.
7. Neville, J.A., Lang, W. and Fleischer, A.B. Jr. (2006) Errors in the Archives of Dermatology and the Journal of the American Academy of Dermatology from January through December 2003. Archives of Dermatology, 142(6), 737–740.

8. Spanou, A., Koletsi, D., Fleming, P.S., Polychronopoulou, A. and Pandis, N. (2016) Statistical analysis in orthodontic journals: are we ignoring confounding? European Journal of Orthodontics, 38(1), 32–38
9. Fleming, P.S., Koletsi, D., Polychronopoulou, A., Eliades, T. and Pandis, N. (2013) Are clustering effects accounted for in statistical analysis in leading dental specialty journals? Journal of Dentistry, 41(3), 265–27.
10. Machin, D., Campbell, M., Fayers, P., Pinol, A. (1997) Sample size tables for clinical studies. 2nd edition. Oxford: Blackwell Science
11. Schulz, KF., Grimes, DA. (2005) Sample size calculations in randomised trials: mandatory and mystical. Lancet 365(9467), 1348–1353.
12. Schulz, KF., Altman, DG., Moher, D.; CONSORT Group (2010a) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Plos Medicine 7(3), e1000251.
13. Schulz, KF., Altman, DG., Moher, D., for the CONSORT Group. (2010b) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Journal of Clinical Epidemiology 63(8), e1–37.
14. Altman, DG. (1980) Statistics and ethics in medical research: III How large a sample? British Medical Journal 281(6251), 1336–1338.
15. John, M. Hoenig & Dennis, M. Heisey. (2001) The Abuse of Power, The American Statistician, 55(1), 19-24.
16. Boutron, I., Ravaud, P. (2018) Misrepresentation and distortion of research in biomedical literature. Proceedings of the National Academy of Sciences USA. 13;115(11), 2613-2619.
17. Farrokhyar, F1., Reddy, D., Poolman, RW. Bhandari, M., (2013) Why perform a priori sample size calculation? Canadian Journal of Surgery. 56(3), 207-13.

18. Charles, P., Giraudeau, B., Dechartres, A., Baron, G., Ravaud, P. (2009) Reporting of sample size calculation in randomised controlled trials: review. British medical journal 338, b1732.
19. Koletsi, D., Fleming, P. S., Seehra, J., Bagos, P. G., & Pandis, N. (2014). Are sample sizes clear and justified in RCTs published in dental journals? Plos one, 9(1), e85949, eCollection 2014.
20. Koletsi, D., Pandis, N., Padhraig, S., Fleming (2014); Sample size in orthodontic randomized controlled trials: are numbers justified? European Journal of Orthodontics, 36(1), 67–73.
21. Koletsi, D., Madahar, A., Fleming, P.S. and Pandis, N. (2015) Statistical testing against baseline was common in dental research. Journal of Clinical Epidemiology, 68(7), 776–781
22. von Elm E., Altman, DG., Egger, M., Pocock, SJ., Gøtzsche, PC., Vandenbroucke, JP. (2007) The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Plos Med 4, e296.
23. Kilkenny, C., Browne, WJ., Cuthill, IC., Emerson, M., Altman, DG. (2012) Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. Osteoarthritis and Cartilage, 20(4), 256–260.
24. Levine, M., Ensom, MH. (2001) Post Hoc Power Analysis: An Idea Whose Time Has Passed? Pharmacotherapy: Official Journal of the American College of Clinical Pharmacy, 21(4), 405–409.
25. Smith, AH., Bates, MN. (1992) Confidence limit analyses should replace power calculations in the interpretation of epidemiologic studies. Epidemiology 3(5), 449–452.
26. Goodman, SN., Berlin, JA. (1994) The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Annals of Internal Medicine 121(3), 200–206.

27. Koletsi, D., Karagianni, A., Pandis, N., Makou, M., Polychronopoulou, A., Eliades, T. (2009) Are studies reporting significant results more likely to be published? American Journal of Orthodontics and Dentofacial Orthopedics 136, 632 e.1-5.

Captions

Table 1. Frequency distribution for the reporting of *a priori* power calculation by article characteristic (n=654).

Table 2. Distribution of statistically significant results and reporting of Confidence Intervals (CIs) across study design.

Table 3. Univariable and multivariable logistic regression with Odds Ratios (ORs) and associated 95% CIs for the effect of a range of article characteristics on reporting of *a priori* power calculation or otherwise (n=654).

Figure 1. Flowchart of study selection.

Figure 2. Predictive margins with 95%CIs for the effect of study design across journals on the reporting of *a priori* power consideration.

Table 1. Frequency distribution for the reporting of *a priori* power calculation by article characteristic (n=654).

	A priori power calculation			**p-value***
	No N (%)	**Yes N (%)**	**Total N (%)**	
Journal				0.22
AJODO	162 (65.9)	84 (34.1)	246(100.0)	
ANGLE	136 (61.3)	86 (38.7)	222 (100.0)	
EJO	87 (62.6)	52 (37.4)	139 (100.0)	
OCR	36 (76.6)	11 (23.4)	47 (100.0)	
Year				0.36
2016	146 (63.8)	83 (36.2)	229 (100.0)	
2017	157 (67.7)	75 (32.3)	232 (100.0)	
2018	118 (61.1)	75 (38.9)	193 (100.0)	
Continent				0.51
America	119 (64.3)	66 (35.9)	185 (100.0)	
Europe	136 (67.3)	66 (32.7)	202 (100.0)	
Asia/other	166 (62.2)	101 (37.8)	267 (100.0)	
No. authors				0.35
1- 3	76 (62.3)	46 (37.7)	122 (100.0)	
4- 6	257 (63.3)	149 (36.7)	406 (100.0)	
≥ 7	88 (69.8)	38 (30.2)	126 (100.0)	
No. centers				0.23

Single-center	141 (61.3)	89 (38.7)	230 (100.0)	
Multi-center	280 (66.0)	144 (34.0)	424 (100.0)	
Study Category				<0.001
Observational	200 (69.4)	88 (30.6)	288 (100.0)	
Interventional	86 (42.0)	119 (58.0)	205 (100.0)	
In-vitro	72 (82.8)	15 (17.2)	87 (100.0)	
Animal	63 (85.1)	11 (14.9)	74 (100.0)	
Significance				0.002
No	94 (54.7)	78 (45.3)	172 (100.0)	
Yes	327 (67.8)	155 (32.2)	482 (100.0)	
Reporting of CIs				0.002
No	359 (67.1)	176 (32.9)	535 (100.0)	
Yes	62 (52.1)	57 (47.9)	119 (100.0)	
Total	421 (64.4)	233 (35.6)	654 (100.0)	

*Pearson chi-square, CIs: confidence intervals

Table 2. Distribution of statistically significant results and reporting of Confidence Intervals (CIs) across study design.

	Study Design					**p-value**
	Observational N (%)	**Interventional N (%)**	**In- vitro N (%)**	**Animal N (%)**	**Total N (%)**	
Significance						0.005*
No	68 (23.6)	68 (33.2)	26 (29.9)	10 (13.5)	172 (26.3)	
Yes	220 (76.4)	137 (66.8)	61 (70.1)	64 (86.5)	482 (73.7)	
Reporting of CIs						<0.001#
No	234 (81.3)	151 (73.7)	80 (92.0)	70 (94.6)	535 (81.8)	
Yes	54 (18.7)	54 (26.3)	7 (8.0)	4 (5.4)	119 (18.2)	
Total	288 (100.0)	205 (100.0)	87 (100.0)	74 (100.0)	654 (100.0)	

*Pearson chi-square, #Fisher's exact test, CIs: confidence intervals

Table 3. Univariable and multivariable logistic regression with Odds Ratios (ORs) and associated 95% Confidence Intervals (CIs) for the effect of a range of article characteristics on reporting of *a priori* power calculation or otherwise (n=654).

Category	**Univariable**			**Multivariable**		
	OR	**95% CI**	**p-value***	**OR**	**95% CI**	**p-value***
Journal			0.23			0.13
AJODO	Reference			Reference		
ANGLE	1.22	0.84, 1.78		1.52	1.01, 2.30	
EJO	1.15	0.75, 1.78		1.32	0.83, 2.12	
OCR	0.59	0.29, 1.22		0.79	0.37, 1.72	
Year			0.37			
2016	Reference					
2017	0.84	0.57, 1.24				
2018	1.12	0.75, 1.66				
Continent			0.51			
Asia/other	Reference					
America	0.91	0.62, 1.35				
Europe	0.80	0.54, 1.17				
No. Authors			0.36			
1-3	Reference					
4-6	0.96	0.63, 1.46				
≥ 7	0.71	0.42, 1.21				
No. Centers			0.23			
Single center	Reference					
Multi center	0.81	0.58, 1.14				
Study Category			<0.001			<0.001
Observational	Reference			Reference		
Interventional	3.14	2.16, 4.57		3.02	2.06, 4.42	

In-vitro	0.47	0.26, 0.87		0.45	0.24, 0.83	
Animal	0.40	0.20, 0.79		0.44	0.22, 0.89	
Significance			0.002			0.04
Yes	Reference			Reference		
No	1.75	1.23, 2.50		1.49	1.01, 2.20	
Reporting of CIs						0.09
No	Reference			Reference		
Yes	1.88	1.25, 2.80	0.002	1.46	0.94, 2.27	

* Wald test for the overall association; CIs: Confidence Interval

Figure 1.

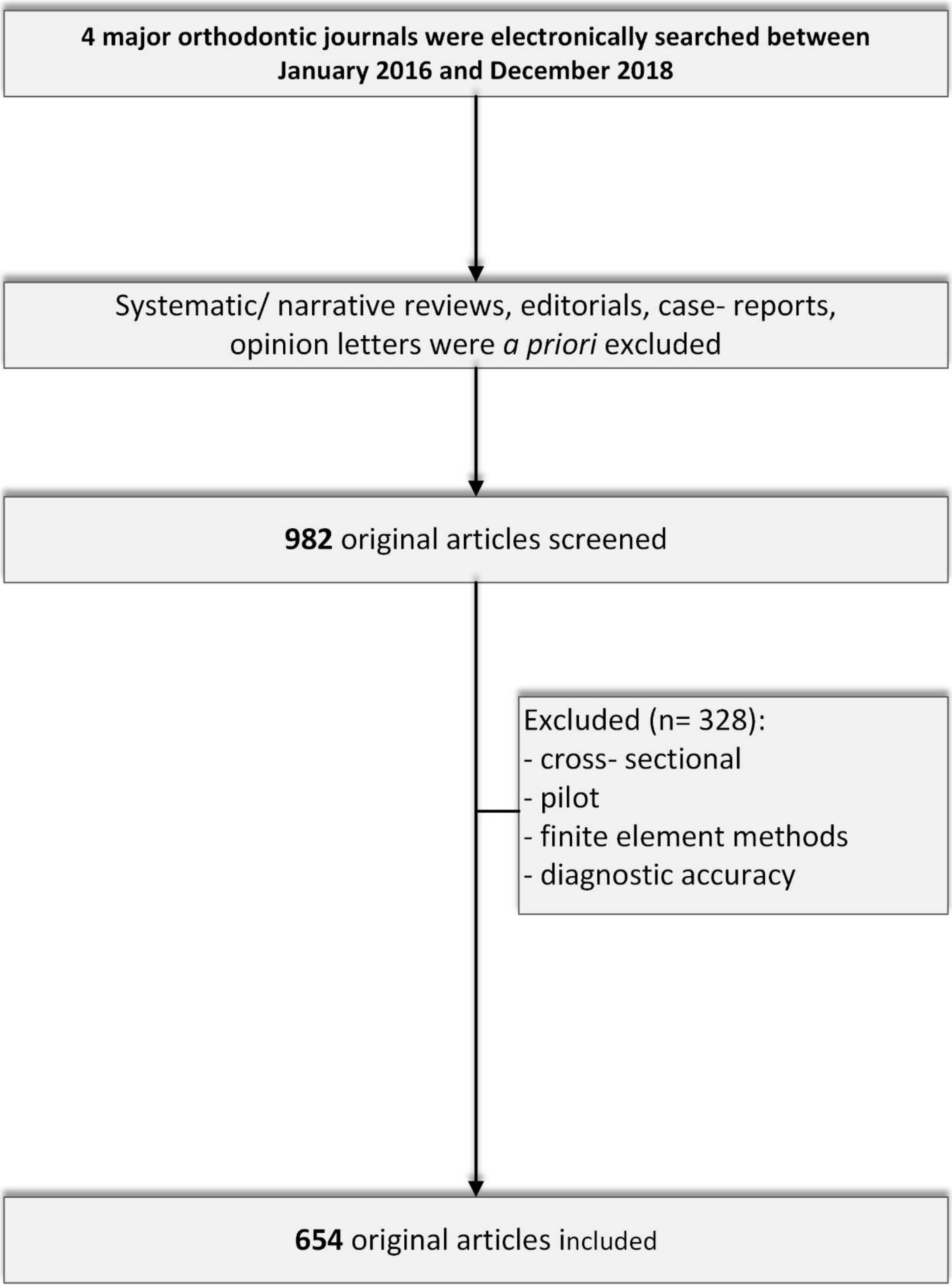
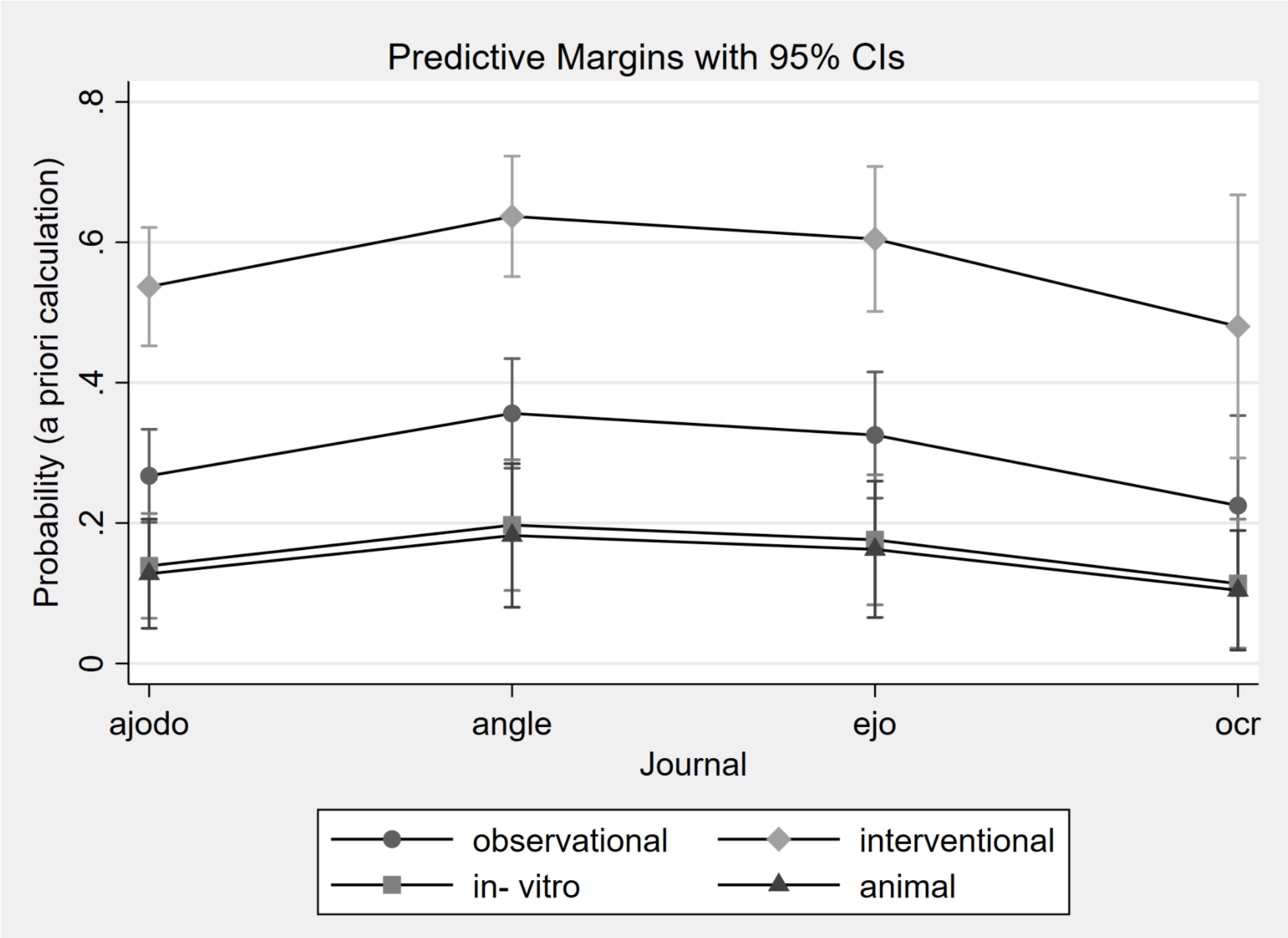

4 major orthodontic journals were electronically searched between January 2016 and December 2018
Systematic/ narrative reviews, editorials, case- reports, opinion letters were *a priori* excluded
982 original articles screened
Excluded (n= 328):
- cross- sectional
- pilot
- finite element methods
- diagnostic accuracy
654 original articles included

Figure 2.



Predictive Margins with 95% CIs
Probability (a priori calculation)
0
.2
.4
.6
.8
ajodo
angle
ejo
ocr
Journal
observational
interventional
in- vitro
animal

Appendix 1. Total number of articles per journal (n=982)

AJODO: 388
ANGLE: 317
EJO: 205
OCR: 72